DOCUMENT RESUME

ED 082 776 LI 004 497

AUTHOR Weeks, Kenneth
TITLE Determination of Pre-Acquisition Predictors of Book Use. Final Report.
INSTITUTION California Univ., Berkeley. Inst. of Library Research.
REPORT NO ILR-73-03
PUB DATE Jul 73
NOTE 25p.; (8 references)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Libraries; *Library Acquisition; *Library Circulation; Library Materials; *Library Material Selection; Use Studies

ABSTRACT

Prediction of book circulation is seen as one major factor for some libraries in the development of an efficient book buying policy. The study approaches this problem of predicting book use from the conviction that there are certain characteristics associated only with high-use books, and certain other characteristics which are associated with little-used books. The object, then, is to identify these high-use, and no-use or low-use indicators so that they may be built into a book selection policy. The suggested test indicators included such things as English language, Major trade publisher, University press, and Conference proceedings. Some of these factors did turn out to be high-performance indicators, but for various reasons, among which may be mentioned the fact that most applied to only a small proportion of the collection sampled, only the English language indicator could have any appreciable impact on the selection process. (Author)

ILR-73-03

Final Report

Determination of Pre-Acquisition Predictors of Book Use

Ken Weeks

Institute of Library Research
University of California

Berkeley, California 94720

July 1973

ABSTRACT

Prediction of book circulation is seen as one major factor for some libraries in the development of an efficient book buying policy. The study approaches this problem of predicting book use from the conviction that there are certain characteristics associated only with high-use books, and certain other characteristics which are associated with little-used books. The object, then, is to identify these high-use, and no-use or low-use indicators so that they may be built into a book selection policy. The suggested test indicators included such things as "English language," "Major trade publisher," "University press," and "Conference proceedings." Some of these factors did turn out to be high-performance indicators, but for various reasons, among which may be mentioned the fact that most applied to only a small proportion of the collection sampled, only the "English language" indicator could have any appreciable impact on the selection process.

TABLE OF CONTENTS

TABLES

FIGURES

I. STATEMENT OF THE PROBLEM AND OBJECTIVES

The inflationary rise of book prices and the decrease in amount or purchasing power of acquisitions budgets force many libraries to look at the problem of how to select the books that represent the greatest value for that library.

This notion of measuring book value is itself problematic. Should one measure the value of a book by the amount of money patrons are willing to pay for it? Or is the amount of use it gets a better measure of a book's value? Then there is the problem of how to measure amount or level of use: what, for example, does frequency of check-out really tell about book use? And how does one measure in-house use? Moreover, there are certain collections (e.g., large academic or research collections) where frequency of check-out, or circulation volume, has little or nothing to do with book value.

Nevertheless, in this imperfect world compromise is often necessary; practical decisions must be made. And if one allows that some notion of book value may be obtained from a study of circulation statistics, or that the amount of use a book gets (as something to be estimated on the basis of its circulation history) can be taken as a measure of its value, then what is needed if book budgets are to be spent in such a way as to bring about the maximization of benefits to the community served by the collection, is some way of predicting circulation before the book is acquired.

This point is appreciated by the State of California Department of Finance report on "Library Cooperation: a Systems Approach to Inter-institutional Resource Utilization" [hereafter referred to as the Auditors' Report], which recommends that in a proposed network of libraries each member should purchase low-use material only in its area(s) of specialization.[1] Clearly, such a notion presupposes a way of effectively predicting book use prior to acquisition.

This, then, was the purpose of this study: to attempt to develop a methodology or set of algorithms for predicting book use in a large university library so that the opportunity to purchase potential high-use items might be compromised as little as possible by the expenditure of funds for other materials.

The major part of the study was done as part of a Design Seminar conducted in the University of California, Berkeley School of Librarianship by Professors R. Swank, M. Cooper, and C. Bourne, with later support provided by the Institute of Library Research.

Specifically, two hypotheses were to be tested: 1) that there are certain indicators which are particularly characteristic of high-use (=most valuable) and no-use (=least valuable) books, and 2) that these indicators could be used, at a stage prior to acquisition, to predict

future book use (=book value).

Before going on to specify what predictors were tested, a brief review of previous work in this area may be helpful.

II. SUMMARY OF PREVIOUS WORK

A recent study by Evans[2] has suggested that books selected by librarians tend to circulate more than those selected by faculty or those received on approval plans, but it fails to identify the methodology or criteria by which librarians do their selecting.

Other studies have approached the problem of predicting book circulation from the standpoint that "reliable conclusions about book usage can be reached by an examination of the characteristics of books themselves rather than of how they were acquired or who selected them."[3] Fussler and Simon, for instance, were concerned with the identification of low-use items in a collection so that, upon their removal to storage, additional space is made available for new acquisitions. Their study showed[4] that a reliable indicator of future use of an item is its past use.

Likewise, Trueswell has applied, in a whole series of articles, the techniques used for managing business inventories--especially the 80/20 rule--to libraries for the purposes of stack-thinning, core collection development, multiple copy determination, and determination of the optimum size of a library's collection. The key statistic for all his strategies is also past use.[5]

But these approaches to circulation-prediction cannot easily be applied to the problem of selection of new books, for which there is no "past use" statistic for that particular library.

In a slightly different vein, McGrath has pointed up a connection between the subject of a book and its level of use: a book tends to circulate if its subject matter corresponds to a profile of the college or university based on courses of instruction offered there. But the effect of such an approach to the problem of book selection would seem to be limited to libraries serving institutions which emphasize a small number of highly specialized academic programs or disciplines; e.g., a mining school, or seminary. 3

Actually, circulation statistics, such as are available from the Loan Department of the Main Library of the University of California, Berkeley, can provide an even clearer picture of use-patterns. The Loan Department at UCB publishes an annual statistical summary which shows the activity of the various parts of the collection, or of the various classes of material (corresponding generally to subject areas).[6] But it is not enough to know that there is nearly twice the activity among items in the B - BJ class (philosophy, logic, metaphysics) as there is in the GN - GV class (anthropology, folklore, games); for although one might, rightly or wrongly, be disposed to spend twice the amount of money for materials in the former subject areas, this still leaves the problem of <u>which</u> materials to purchase, in either area.

III. SCOPE OF THE STUDY: POPULATION AND CONSTRAINTS

Because this study, in a sense, had been spawned by the Auditors' Report, it seemed appropriate to adopt their definitions and constraints where possible. Therefore, the universe we sampled consisted, like theirs, mostly of monographs and monographic serials; periodicals, theses, phonograph recordings, maps, and art prints were excluded. Although the Auditors also attempted to exclude obvious gifts, no such attempt was made here. Furthermore we used the same algorithm devised by the Auditors and used by them in their report to categorize the level of use of each item in our own sample (Figure 1).

There were, however, some constraints peculiar to our study. To begin, whereas the Auditors drew their sample from the collections of six California State University and College libraries, our sample was drawn entirely from the Loan Stack Collection of the Main Library of the University of California at Berkeley.

Second, the population we sampled consisted, to be precise, not of all monographs and monographic serials, but rather, of those monographs and monographic serials which were on the shelves at the time of the data collection. (For more on this, see the section on SAMPLING TECHNIQUE.) Put briefly, our sample was therefore biased against the set of books not on the shelves at that time--which most likely means biased against high-use items. Probability of bias is estimated at a maximum of 8.6%, corresponding to the largest percent of the collection in circulation at any one time [i.e., (estimated total number of volumes in circulation) divided by (total number of volumes in collection)].*

Third, while we did use the Auditors' algorithm to categorize level of use, nevertheless this categorization was made only on the evidence of extant transaction records. The circulation history of each book is recorded on date-due slips pasted on the end paper. When a date-due slip is removed, either by accident or design, the complete record does not survive. In fact, this happens often, though no statistics were kept on the number of occurrences within our sample. The point, however, is that when in our study part or all of the record might have perished, the algorithm for categorizing the level of book use was strictly applied to the surviving data only, if any; no estimates were attempted, though

---

* 90,100/1,049,487. I am indebted to Mr. Tim DeWolf of the Loan Department, Main Library, UCB, for this figure, which is based on total holdings of all Loan Department books, including those in Richmond storage. It does not include any items not recorded in the loan files, such as department charges and books waiting to be reshelved.

it might have seemed reasonable to automatically rank as high-use an item from which, say, two date-due slips had been removed. On page 25 of their Report, the Auditors acknowledge this problem of the risk of measurement error. They suggest, however, on the basis of a measurement of actual error that they were able to make, that in fact it amounted to only a small risk for their sample data. For us who were not able to obtain a measurement of actual error, the risk amounts at worst to the fact that we may have applied, to some extent, a yet more conservative measure of book use; i.e., the number of low-use items grows even larger, while the number of high-use items decreases. This is a significant point because the result is a further refinement of the high-use group and perhaps yields a purer set of high-use predictors.

Fourth, the "no-use" category doubtless includes a certain percentage of books only recently cataloged and, therefore, available to patrons for only a short time; even if one supposes that use equals value, it is a bit rigid to say that all "no-use" items are the books of least value. Actually, this must be true of the data collected by the Auditors themselves, but they do not make a point of it in their Report. The presence of this subset of "no-use/potential high-use (or no-use)" items could complicate matters by introducing what might be uncharacteristic predictors into the "no-use" bank. No statistics were kept on the number of sample items that fell in this category, but of the total number of monographs in the collection as of June 30, 1972, approximately 2.3% had been added within the fiscal year ending on that date. *

Fifth, we did not consider interlibrary loan transactions, although they are so indicated on the date-due slips, in computing level of book use. Our reasoning was that level of book use was properly defined by local use. No record was kept of the number of ILL transactions we so ignored.

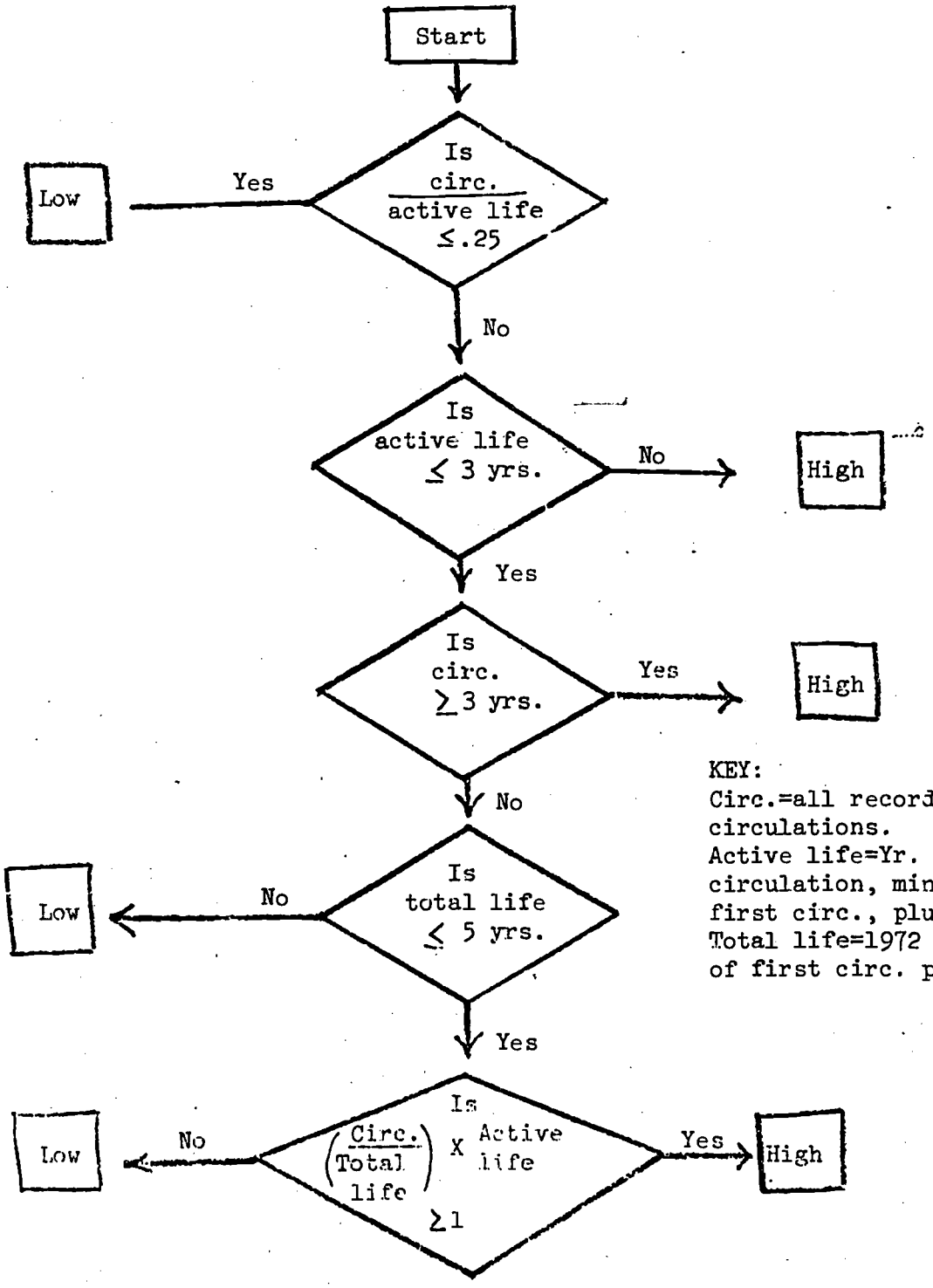
Finally, unlike the Auditors, we did not correct for unrecorded or in-house usage. The effect of this, to judge from the Auditors' analysis on pages 36-40 of their Report, was a failure to correctly identify a certain percentage of "'true' high-use" and "'true' low-use" items, having misplaced them in a lower use category.

---

* This percentage is based on figures for total "bound volumes" in the "main i.e., excluding Bancroft, Morrison, and all branch libraries. From the "Size of the Libraries of the University of California. Table II: Bound olumes and Current Serials (Berkeley Details), 27:35 (21 September 1972), p. 10.

Figure 1

The Auditors' Book Classification Algorithm



KEY:
Circ.=all recorded circulations.
Active life=Yr. of last circulation, minus year of first circ., plus one.
Total life=1972 minus year of first circ. plus one.

Source: Auditors' Report

IV. TEST PREDICTORS

In order to qualify for testing, a "predictor" or "indicator" had to satisfy the basic requirement that it be among those of which a bibliographer or selections officer might reasonably be expected to have some knowledge from his reading of pre-publication notices and advertisements or such other usual acquisition tools as library or publishers' catalogs. In addition, hunch, intuition, and examination of a preliminary sample of 50 items played a part in this selection process. This method eventually yielded the following list of original test indicators:

1. English language: In the case of bi- or multi-lingual books, any item of which a substantial part of the text was in English was regarded as an English-language book.

2. Single author--pesonal: Anonymous or pseudonymous works were not included.

3. Multiple author--personal: Edited works were included, but translators were not regarded as contributing to multiplicity of authorship.

4. Major trade publisher: Originally this was tentatively and arbitrarily taken to include only the following American publishers: McGraw-Hill, Wiley, Prentice-Hall, Crowell/Collier/MacMillan, College & University Press, Norton, Van Nostrand, and Johnson Reprint Company. Later the list was expanded to include the following: Viking, Dutton, Harper, Putnam, Sheed & Ward, Dodd Mead, Lippincott, Holt, Scribner's, Knopf, Houghton Mifflin, and Appleton. For further details see the section on TEST RESULTS.

5. University press: The sample was not restricted to American university presses.

6. Bibliography, catalog, abstract, annual review, and conference proceedings. "Conference proceedings" was later added as a separate indicator. For further details see the section on TEST RESULTS, part C.

7. Handbook, manual guide.

8. Illustrated: Photographs, drawings, etc.; these were to be distinguished from the types of graphic materials listed under indicator 9.

9. Maps, charts, diagrams, tables.

10. Work complete in this volume.

"Price" was not tested because this information was not available for the test either fr[illegible]e public shelf list or from the book itself. Nor was "Imprint date" [illegible]ed, although a significant percentage of books selected for academic libraries are non-current items. For example, according to statistics for the libraries of the nine University of California campuses, 37% of all monographic materials cataloged during the period 1963 to 1967 bore a publication date of 1949 or earlier.[8] It was nevertheless felt that these non-current items come especially recommended by faculty or staff, in which case some of the uncertainty about future level of use is already settled.

V. SAMPLING TECHNIQUE AND DATA COLLECTION

A. SAMPLE I

To begin, a sample of 50 items was taken from the public shelf list. The shelf list was used as the data base because, unlike the author/title and subject catalogs, 1) each item is entered only once and all items thus have an equal chance of being selected, and 2) the shelf list contains only items found in the Loan Stack Collection.

Using a random number table, 50 drawers were chosen at random, and a card one-inch from the front of each drawer was noted. When 50 cards had been so selected, the books were retrieved from the shelves. Because of time constraints, no attempts were made to recall or search any items not on the shelves; instead, additional items were selected from the shelf list and retrieved. If a book was not on the shelf, another card was selected from the same drawer, this time one-inch from the back. If this second selection was not on the shelf, a third card was selected from the same drawer, this time two inches from the front, and so on. Eventually, 50 sample items were so collected and classified according to level of use (high, low, or no), using the Auditors' algorithm. The "high-use" and "no-use" groups were then examined for promising predictors of book use. The data breakdown is given in the section on TEST RESULTS; it shows that four predictors appeared to be promising at this stage. An indicator was regarded as promising if the ratios, based on normalized scores, of high-use to no-use, and no-use to high-use were markedly different. The promising predictors, expressed in high-use terms were the following: "English language," "Major trade publisher," "University press," and "Bibliography, catalog, abstract, annual review, conference proceedings."

B. SAMPLE II

A second sample was then obtained used a less time-consuming method. In the first sample no significant effort was made to avoid prejudicing the data collection against books not on the shelves. This was taken one step further in the second sample: the shelf-list was by-passed altogether, and the following rule was used in selecting this next batch of sample items: For each stack level, starting in the first aisle at the north-east end of each row of stacks, and afterwards every other aisle, select the 10th book on the 4th shelf of the 2nd section facing east. This resulted in the collection of 221 items. A log was kept which recorded each item's call number, use-classification, and the presence or absence of the four predictors identified for further testing by the first sample. This sample yielded results which led to testing of additional predictors in the subsequent sample.

C. SAMPLE III

Finally, a third sample was drawn according to a rule similar to that used for collecting the second batch: for each stack level, starting

in the 2nd aisle at the northeast end of each row of stacks, and afterwards every other aisle, select the 5th book on the 3rd shelf of the 3rd section looking west. This yielded 192 sample items, and brought the total sample to 493 items. A log was kept to record the presence or absence of the original ten indicators of the first sample, plus one additional indicator. (For details see the following section on TEST RESULTS.)

VI. TEST RESULTS

A. SAMPLE I

The first sample (n=50) turned up 21 high-use items, 10 no-use items, and 19 low-use items. The test results are summarized in Table 1. Because the frequencies were extremely disparate, particularly between the high-use and no-use groups, the scores for the no-use group were normalized; that is, they were scaled upward as if there were as many no-use books as high-use books in the sample. This facilitated meaningful comparison between the two groups.

These points of comparisons reflect the fact that it was upon the two levels high-use and no-use that our interest centered. "Low-use" was expected to be a slush category which would supposedly contain a mixture of predictors of high-use and predictors of no-use.

Of the 21 high-use items, 16 (76%) were English-language materials; and of the 10 no-use items, 9 (90%) were non-English language. The most important figures were those concerned with a given indicator's comparison of the ratio of its high-use to no-use with the ratio of its no-use to high use. For the English language items, the ratio of high-use to no-use was nearly 8:1, while the ratio of no-use to high-use was 13:100. (Both ratios were based on the normalized no-use figure.) Likewise, for non-English language items the ratio of high-use to no-use was 26:100 (or 0.26), while the ratio of no-use to high-use was 3.78. From this disparity between the two ratios, it appeared that "English language" might be a predictor of high book use and that "non-English language" might predict no-use. This early indication was confirmed throughout our sampling.

In contrast, the other indicators in the first sample appeared to have little or no predictive value. In only two other cases--"Single author--personal" and "Work complete in this volume"--did a predictor show up in significant numbers; but the instances of its presence were evenly divided among the high-use and no-use categories. (See column of normalized scores, Table 1.) More often the case was one of low incidence of a predictor's presence, with the instances of its absence evenly divided between the high-use and no-use levels.

Within this negative context there were three indicators ("Major trade publisher," "University press," and "Bibliography, etc.") where the difference between the ratio of high-use to no-use and the ratio of no-use to high-use seemed large enough to support some prospect of predictive value.

At this stage of data collection, the "Major trade publisher" included the following: McGraw-Hill, Wiley, Prentice-Hall, Crowell/Collier/MacMillan, College & University Press, Norton, Van Nostrand,

Table 1

Frequency Distribution of Test Indicators for Sample I (n_1=50)

Indicators	Number of books with this frequency of use			Normalized scores		Ratios based on normalized scores	
	high-use	no-use	low-use	high-use	no-use	high:no	no:high
English language	16	1	8	16	2.1	7.62	.13
non-English lang.	5	9	11	5	18.9	.26	3.78
Single author--personal	15	6	15	15	12.6	1.19	.84
non-single auth--pers	6	4	4	6	8.4	.71	1.40
Multiple author--personal	5	2	0	5	4.2	1.19	.84
non-mult auth --pers	16	8	19	16	16.8	.95	1.05
Major trade publisher	3	0	2	3	0		
non-mtp	18	10	17	18	21	.86	1.17
University press	4	0	1	4	0		
non-up	17	10	18	17	21	.81	1.24
Bibliography, catalog, abstract, annual review, conf. proceedings	0	2	1	0	4.2		
non-bibliography. . . etc.	21	8	18	21	16.8	1.25	.80
Handbook, manual, guide	3	1	5	3	2.1	1.43	.70
non-handbook. . . etc.	18	9	14	18	18.9	.95	1.05
Illustrated	5	3	9	5	6.3	.79	1.26
not illustrated	16	7	10	16	14.7	1.09	.92
Maps, charts, diagrams	7	5	9	7	10.5	.67	1.50
non-maps. . . etc.	14	5	10	14	10.5	1.33	.75
Work complete this vol.	19	9	17	19	18.9	1.01	.99
Work not complete. . .etc.	2	1	2	2	2.1	.95	1.05

and Johnson Reprint Company. This indicator showed up only 5 times among the 31 high-use and no-use items; of those 5 instances, 4 were high-use. A similar pattern of meager returns yet promising ratios was repeated for "University Press" (4 high-use:1 no-use) and for "Bibliography, etc." (2 no-use:0 high-use). We next tested these 4 indicators in the second sample of 221 items.

B. SAMPLE II

The second sample yielded 63 high-use and 87 no-use items. The striking feature of this sample was that "English language" continued to show as a strong high-use indicator. Table 2 contains the breakdown of these 221 sample items by indicator and use-category. Concerning the English language items, the ratio of high-use to no-use is nearly 4:1, while the ratio of no-use to high-use is about 1:4. (These ratios were based on a normalized score for the high-use group.)

"Non-English language" was also an interesting indicator. Half of the non-English language items were no-use; about 32% were low-use, and only 18% were high-use. The implication is that if non-English language were used as a basis for rejection, only 18 out of 100 books so rejected would have represented eventual high-use items lost to the collection.

Results for the other indicators tested in this sample were disappointing insofar as there were only negligible differences between the high-use:no-use and the no-use:high-use ratios.

Two additional points deserve comment. 1) Clearly, the first sample of 50 was too small to reliably indicate which "predictors" might work. 2) The low incidence of some indicators pointed out that even if such an indicator as "Major trade publisher" were found by itself to be a high-performance indicator, nevertheless it would be of limited value to the bibliographer if books from major trade publishers constituted only a small percentage of the annual acquisitions.

C. SAMPLE III

We then tested a third sample of 192 items for a frequency distribution of the original ten indicators, plus one other indicator; the indicator "Bibliography, catalog, abstract, annual review, and conference proceedings" was divided into two separate indicators: "Bibliography, catalog, abstract, annual review" and "Conference proceedings." We also expanded the number of "Major trade publishers" to include the following: Viking, Dutton, Harper, Putnam, Sheed & Ward, Dodd Mead, Lippincott, Holt, Scribner's, Knopf, Houghton Mifflin, and Appleton.

The third sample yielded 62 high-use items, 69 no-use items, and 61 low-use items. The results are given in Table 3 and are consistent with our earlier findings. Note, for example, that "English language" continues to indicate high-use, while "non-English language" is only slightly less certain as a predictor of no-use. Also, "Major trade publisher,"

Table 2

Frequency Distribution of Test Indicators for Sample II (n_2=221)

Indicators	Number of books with this frequency of use			Normalized scores		Ratios based on normalized scores	
	high-use	no-use	low-use	high-use	no-use	high:no	no:high
English language	36	13	24	49.7	13	3.82	.26
non-English lang	27	74	47	37.3	74	.50	1.98
Major trade publisher	1	0	2	1.4	0		
non-mtp	62	87	69	86.8	87	1.0	1.0
University press	3	4	4	4.2	4	1.05	.95
non-up	60	83	67	84	83	1.01	.99
Bibliography. . . etc.	4	8	5	5.6	8	.7	1.43
non-bibliography. . .etc.	59	79	66	82.6	79	1.05	.96

understood here in its expanded sense, looks very good as a predictor of high-use. However, "non-major trade publisher" was not a predictor of no-use. None of the other frequencies is remarkable.

D. COMBINED SAMPLE

The combined statistics for the three samples are summarized in Table 4. Important to note is that the frequencies for the predictor "Major trade publisher" are not homogeneous because that predictor was defined differently for the third sample. Also, separate statistics for "Conference proceedings" were collected only for the third sample. Finally, Table 4 also includes estimated percentages of the Loan Stack Collection represented by each of the predictors. These estimated percentages were based on a random sample, where n=300, drawn from the public shelf list with the help of a random number table.

E. COMBINATIONS OF INDICATORS

The performance of our test indicators can be improved by taking the indicators in certain combinations. Table 5 displays some interesting combinations. Table 5 also shows that while combining predictors will improve their level of performance, combining them also reduces their impact. For example, "English language" taken with "Major trade publisher" yields a high-use to no-use ratio of 17:1, but represents only 6% of the collection.

Table 3

Frequency Distribution of Test Indicators for Sample III (n_3=192)

Indicators	Number of books with this frequency of use			Normalized scores		Ratios based on normalized scores	
	high-use	no-use	low-use	high-use	no-use	high:no	no:high
English language	42	14	24	46.2	14	3.3	.30
non-English lang	20	55	37	22	55	.4	2.50
Single author--personal	37	45	42	40.7	45	.9	1.11
non-single auth--pers	25	24	19	27.5	24	1.15	.80
Multiple author--personal	21	11	15	23.1	11	2.1	.48
non-mult. auth--pers	41	58	46	45.1	58	.78	1.29
Major trade publisher	11	1	5	12.1	1	12.1	.08
non-mtp	51	68	56	56.1	68	.82	1.18
University press	6	2	3	6.6	2	3.3	.30
non-up	56	67	58	61.6	67	.92	1.09
Bibliography, catalog, abstract, annual review	2	4	0	2.2	4	.55	1.82
non-bibliography. . . etc.	60	65	61	66	65	1.02	.98
Conference proceedings	0	0	2	0	0		
non-conf. proc.	62	69	59	68.2	69	.99	1.01
Handbook, manual, guide	1	2	2	1.1	2	.55	1.82
non-handbook. . . etc.	61	67	59	67.1	67	1.	1.
Illustrated	22	20	17	24.2	20	1.2	.83
not-illustrated	40	49	44	44	49	.89	1.11
Maps, charts, diagrams, tables	18	13	11	19.8	13	1.52	.66
non-maps. . . etc.	44	56	50	48.4	56	.86	1.16
Work complete this vol.	45	45	42	49.5	45	1.1	.91
Work not complete. . . etc.	17	24	19	18.7	24	.78	1.28

Table 4

Frequency Distribution of Test Indicators for Combined Samples (n=463)

Indicators	Number of books with this frequency of use			Normalized scores		Ratios based on Normalized Scores		Est. percent of Collection
	high-use	no-use	low-use	high-use	no-use	high:no	no:high	
English language	94	28	56	107.2	28	3.83	.26	43
non-English lang	52	138	95	59.3	138	.43	2.33	57
Single author--personal	52	51	47	52	53.6	.97	1.03	64
non-single auth--pers	31	28	23	31	29.4	1.05	.95	36
Multiple author--personal	26	13	15	26	13.6	1.91	.52	13
non-multiple auth--pers	57	66	65	57	69.3	.82	1.22	87
Major trade publisher	15	1	9	17.1	1	17.10	.06	6
non-MTP	131	165	142	149.3	165	.90	1.11	94
University press	13	6	8	14.8	6	2.46	.41	8
non-up	133	160	143	151.6	160	.95	1.06	92
Bibliography, catalog, abstract, annual review, conference proceedings	4	10	6	4.6	10	.46	2.17	12
non-biblio. . . etc.	80	87	84	92	87	1.06	.95	88
Conference proceedings	0	0	2	0	0			1
non-conf. proc.	62	69	59	68.2	69	.99	1.01	99
Handbook, manual, guide	4	3	7	4	3.2	1.25	.8	4
non-handbook. . . etc.	79	76	73	79	87.4	.90	1.11	96
Illustrated	27	23	26	27	24.2	1.12	.90	36
not illustrated	56	56	54	56	58.8	.95	1.05	64
Maps, charts,. . . tables	25	18	20	25	18.9	1.32	.76	18
non-maps. . . etc.	58	61	60	58	64	.91	1.10	82
Work complete this vol.	64	54	59	64	56.7	1.13	.89	77
Work not complete, etc.	19	25	21	19	26.2	.73	1.38	23

Note that statistics were not kept for all indicators for all samples.

Table 5

Frequency Distribution of Combinations of Indicators for All Samples

Indicators	Number of books with this frequency of use			Normalized scores		Ratios based on Normalized scores		Est. percent of collection
	high-use	no-use	low-use	high-use	no-use	high:no	no: high	
English language plus Major trade publisher	16	1	8	18.24	1	18.24	.05	6
English language plus Single author--personal	38	6	18	38	6.3	6.03	.17	26
English language plus Single author--pers plus Major trade publisher	11	1	6	11	1.05	10.47	.1	5
English language plus University Press	12	2	7	13.68	2	6.84	.15	5
English language plus Single auth--pers plus University press	7	0	4	7	0			4
English language plus Multiple author--pers	20	3	8	20	3.15	6.35	.16	7
English language plus Multiple author--personal plus Work complete this volume	16	1	3	16	1.05	.07		6

* Expanded sense

VII. CONCLUSIONS

Our data collection and analysis showed that there was, indeed, a correlation between "English language" and high-use, and between "non-English language" and no-use. However, even if one supposes that English-language materials represent a good investment of book funds, there still remains the problem of deciding which English language items to acquire, as no library book budgets are sufficiently large to cover the purchase of all English-language items.

One must also consider that none of these predictors is infallible. This is suggested by the case of "English language," a moderately high-performance predictor that represents a large part of the collection. If one were to purchase all 178 English-language items included in our combined sample, the result would be the acquisition of 28 new items (about 16%) which might never be used, or 84 items (47%) which may never be high use.

We also note a serious methodological flaw. 94 of 178 English language items in our combined sample turned out to be high-use. This does not give reason to suppose that the total population of all English-language books would yield the same percentage of high-use items. For our sample does not represent that universe of all English-language books, but rather, a universe (large university research collection) where pre-selection has already taken place, although perhaps not on the basis of anticipated use. This point is equally applicable to all our test indicators and further reduces by an unknown factor whatever value they might be thought to possess in the book-selection.

In summary, we found four potentially good indicators of high-use. In decreasing order of performance level, they were as follows: "Major trade publisher," "English language," "University press," and "Multiple author--personal." Two good indicators of no-use from the many tested indicators were "non-English language" and "Bibliography, catalog, abstract, annual review, conference proceedings." However, we do not know how effective these indicators would be for any material beyond that which has been pre-selected by the library selections staff. Furthermore, if four of these six potentially good indicators were applied ("English language" and "non-English language" excepted), they would correspond to only a small fraction (6 to 13%) of the collection as it is presently constituted. Unfortunately, there is no means of predicting the percentage of annual acquisitions (a more meaningful category in the context of this report) to which these indicators correspond.

REFERENCES

Note No.

1 California Department of Finance, Audits Division and Budget Division, "Library Cooperation: A Systems Approach to Interinstitutional Resource Utilization." Burton R. Cohen, Audits Division Project Director (unpublished draft report, 1972).

2 Evans, G. Edward, "Book Selection and Book Collection Usage in Academic Libraries," Library Quarterly, 40 (1970) pp. 297-308.

3 McGrath, William E., "The Significance of Books Used according to a Classified Profile of Academic Departments," College and Research Libraries, 33:3 (May 1972) pp. 212-219.

4 Fussler, Herman E., and Julian L. Simon, Patterns in the Use of Books in Large Research Libraries (Chicago: University of Chicago Press, 1969).

5 Trueswell, Richard W., "Determining the Optimal Number of Volumes for a Library's Core Collection," Libri, 16 (1966) pp. 49-60.

_____, "A Quantitative Measure of User Circulation Requirements and Its Possible Effect on Stack Thinning and Multiple Copy Determination," American Documentation, 16 (1965) pp. 20-25.

_____, "Some Behavioral Patterns of Library Users: The 80/20 Rule," Wilson Library Bulletin, 43:5 (January 1969) pp. 458-461.

_____, "User Circulation Satisfaction vs. Size of Holdings at Three Academic Libraries," College and Research Libraries, 30:3 (May 1969) pp. 204-13.

6 [University of California, Berkeley: General Library], Current Holdings in the Loan Department (Berkeley: University of California, 1972) 2nd ed., August 1972.